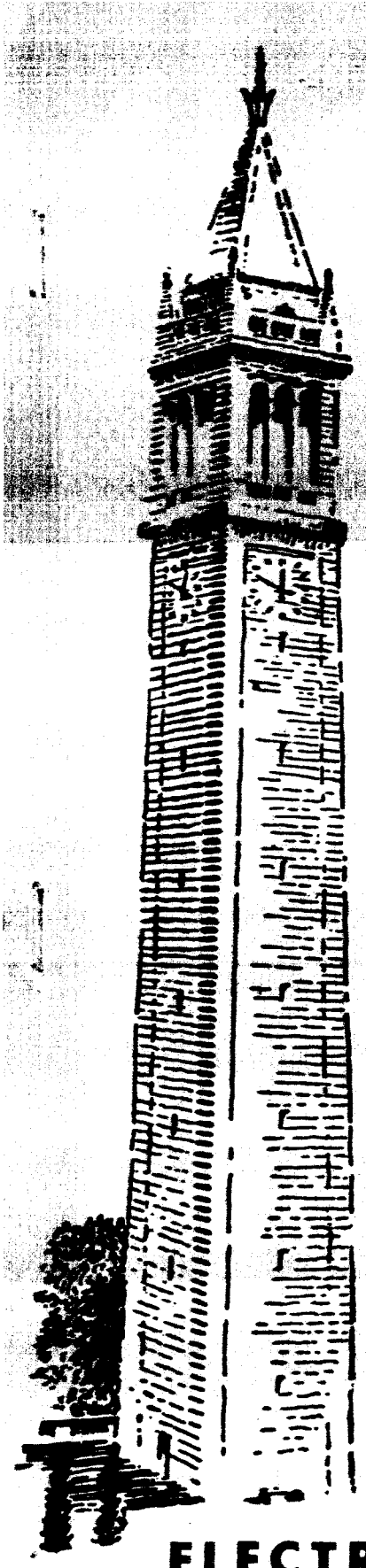

UNPUBLISHED PRELIMINARY DATA

Internal Technical Memorandum M-94

A NONLINEAR RECURSIVE ESTIMATION PROCEDURE AND ITS APPLICATION TO CERTAIN NAVAGATION PROBLEMS

by

D. J. Sakrison

August 18, 1964

ELECTRONICS RESEARCH LABORATORY

UNIVERSITY OF CALIFORNIA

BERKELEY, CALIFORNIA

Electronics Research Laboratory
University of California
Berkeley, California
Internal Technical Memorandum M-94

A NONLINEAR RECURSIVE ESTIMATION PROCEDURE AND ITS APPLICATION TO CERTAIN NAVAGATION PROBLEMS

by

D. J. Sakrison

The research herein was supported by National Aeronautics and Space Administration under Grant NsG-354 (S-1).

August 18, 1964

A Nonlinear Recursive Estimation Procedure and Its Application to Certain Navigation Problems*

D. J. SAKRISON**

ABSTRACT

33171

In the past ten years, a class of recursive estimation methods, referred to as Stochastic Approximation methods, has been developed to handle problems where the observations without errors are nonlinear functions of the parameters to be estimated [1-3]. Recently, such a method has been developed to yield asymptotically efficient estimates in the type of estimation problems encountered in radar and radio astronomy [4]. The basic method used is also applicable to some of the estimation problems arising in space navigation. Our purpose here is to apply this method to navigation and carry out the required modifications in the analysis. This method has the following interesting features: (1) it is recursive and thus it allows real time computation without extensive computing facilities; (2) it is applicable to a broad class of problems where the observables are nonlinear functions of the parameters to be estimated.

*The research herein was supported by National Aeronautics and Space Administration under Grant NsG-354 (S-1).

**The author is affiliated with the Department of Electrical Engineering and Electronics Research Laboratory, University of California, Berkeley, California.

A Nonlinear Recursive Estimation Procedure and Its Application to Certain Navigation Problems*

D. J. SAKRISON**

I. INTRODUCTION

We consider here problems best typified by the problem of trying to estimate the six orbital elements of a satellite or space vehicle in orbit. We assume that the orbit is sufficiently high that, for the number of orbits over which observations are to be carried out, atmospheric drag can be neglected and perturbations due to departures of the gravitational field from shperical are minor enough that they can be satisfactorily computed from the nominal orbit. Thus, during the observation interval, the orbit can be satisfactorily represented by an ellipse plus a known gravitational perturbating term. Our objective is to estimate the six elements of this ellipse from a sequence of observations such as ground radar sightings, on board radar beacon sightings, star tracker observations, etc.

The method we consider here has the following advantages:

(1) since it is recursive, it allows real time computation and requires only minimal computer storage; it also requires inversion of only low-order matrices;

(2) it directly and exactly handles the case in which the observations are nonlinear functions of the parameters to be estimated.

This method has the disadvantage of being useful only when the number of observations made is large, for the error associated

*The research herein was supported by National Aeronautics and Space Administration under Grant NsG-354 (S-1).

**The author is affiliated with the Department of Electrical Engineering and Electronics Research Laboratory, University of California, Berkeley, California.

with the estimate computed at the end of a small number of observations is unduly large. However, under certain conditions, this method is asymptotically efficient; that is, the estimation error approaches the bound given by the Cramér-Rao inequality as the number of observations becomes large. Thus, we will focus our analysis on the asymptotic behavior of the estimate.

Let the observations be made at times $t = 1, 2, 3, \ldots,$ and denote the reading of the i-th of Q instruments at time t by

$$z_t^i = f_t^i(\underset{\sim}{v}) + b^i + e_t^i \qquad i = 1, 2, \ldots, Q \qquad t = 1, 2, 3\ldots$$

in which $f_t^i(\underset{\sim}{v})$ is a known function of the orbital elements (denoted by the six-dimensional vector $\underset{\sim}{v}$); b^i is the bias associated with the i-th instrument, and e_t^i is a zero-mean gaussian random variable whose variance depends on the instrument and also possibly on the time t. We assume that observations made at different times, either from the same or different instruments, are uncorrelated.

In addition to the orbital elements, $\underset{\sim}{v}$, we may wish to estimate some or all of the bias terms. Let the total number of quantities to be estimated be R, $6 \leq R \leq 6 + Q$. We will denote an arbitrary set of values of these R quantities by the R-dimensional vector $\underset{\sim}{x}$ and the true value of these quantities by the vector $\underset{\sim}{\theta}$.

For reasons that will be apparent later, we will group together the QK observations made from all Q instruments at the times

$$t = nK + 1, \quad nK + 2, \ldots, \quad (n + 1)K.$$

We pick K to be the smallest integer such that the mean values of these QK measurements uniquely determine $\underset{\sim}{x}$. (If not all the

bias terms are included in $\underset{\sim}{x}$, this unique determination will be in error). We denote this collection of observations made from time $t = nK$ to $t = (n + 1)K$ by the QK dimensional vector $\underset{\sim}{z}_n = \underset{\sim}{f}_n(\underset{\sim}{\theta}) + \underset{\sim}{e}_n$, in which all the bias terms have been included in $\underset{\sim}{f}_n(\underset{\sim}{x})$: if $\underset{\sim}{x}$ does not include all the bias terms, our final estimate will contain an error determined directly by the bias terms not included in $\underset{\sim}{x}$. Thus $\underset{\sim}{e}_n$ is a zero-mean gaussian vector random variable which is uncorrelated for different values of n and whose covariance matrix we denote by

$$E\left\{\underset{\sim}{e}_n' \underset{\sim}{e}_n\right\} = \underset{\sim}{\Sigma}_n ,$$

the prime denoting the transpose of a vector (or matrix). If we assume the errors from different instruments are uncorrelated, then $\underset{\sim}{\Sigma}_n$ will be diagonal. Our uniqueness requirement implies that the function $\underset{\sim}{f}_n(\underset{\sim}{x})$ possesses a unique inverse for all values of $\underset{\sim}{x}$ which are regarded a priori as possible.

Having introduced the necessary notation, we proceed as follows. In Sec. II we find the bound given by the Cramèr-Rao inequality for the minimum error achievable by an unbiased estimator. In Sec. III we describe the recursive estimation method, develop an expression for the asymptotic behavior of the error in the estimates generated by this method, and relate this to the bound given by the Cramèr-Rao inequality.

II. THE CRAMÈR-RAO BOUND ON THE ESTIMATION ERROR

We wish to apply the bound of the Cramèr-Rao inequality to the situation described in Sec. I. Although simple proofs of this inequality are widely available in the one-dimensional case [5], this is not true for the multidimensional case. Therefore, we first give a short derivation of this inequality.

Let the components of the vector $\underset{\sim}{z}$ denote all the observations made and let the distribution of $\underset{\sim}{z}$ for a value $\underset{\sim}{x}$ of the vector to be estimated be denoted by

$$dP(\underset{\sim}{z}\,;\,\underset{\sim}{x}) = p(\underset{\sim}{z}\,:,\,\underset{\sim}{x})\,d\underset{\sim}{z}.$$

Let $\hat{\underset{\sim}{x}} = \hat{\underset{\sim}{x}}(\underset{\sim}{z})$ denote an arbitrary unbiased estimator of $\underset{\sim}{\theta}$. Using the fact that $\hat{\underset{\sim}{x}}$ is unbiased, we have for any set of constants a_j

$$\sum_{j=1}^{R} a_j \int (\hat{x}_j - \theta_j) dP = 0 \,. \tag{1}$$

Taking the partial derivative of both sides of this equation with respect to x_i

$$a_i = a_i \int dP = \sum_{j=1}^{R} a_j \int (\hat{x}_j - \theta_j) \frac{\partial}{\partial x_i}\, dP$$

$$= \Sigma a_j \int (\hat{x}_j - \theta_j) \left(\frac{\partial}{\partial x_i}\, p\right) dz$$

$$= \Sigma a_j \int (\hat{x}_j - \theta_j) \left(\frac{\partial}{\partial x_i} \ln p\right) p\, dz$$

$$= \Sigma a_j \int (\hat{x}_j - \theta_j) \left(\frac{\partial}{\partial x_i} \ln p\right) dP. \tag{2}$$

Let us multiply both sides of this equation by an arbitrary constant b_i and then sum over i the equations that result for all R values of i. Denoting the inner product (scalar product or dot product) of the vectors $\underset{\sim}{a}$ and $\underset{\sim}{b}$ by $(\underset{\sim}{a}, \underset{\sim}{b})$, we obtain

$$(\underset{\sim}{a}, \underset{\sim}{b}) = \int \left[\sum_{j=1}^{R} a_j (\hat{x}_j - \theta_j) \right] \left[\sum_{i=1}^{R} b_i \left(\frac{\partial}{\partial x_i} \ln p \right) \right] dP. \tag{3}$$

Let

$$\underset{\sim}{X} = E \left\{ (\hat{\underset{\sim}{x}} - \underset{\sim}{\theta})' (\hat{\underset{\sim}{x}} - \underset{\sim}{\theta}) \right\} \tag{4}$$

denote the covariance matrix of the estimation errors and let $\underset{\sim}{B}$ denote the covariance matrix whose i-j-th element is

$$b_{ij} = E \left(\frac{\partial}{\partial x_i} \ln p \right) \Big|_{\underset{\sim}{x}=\underset{\sim}{\theta}} \left(\frac{\partial}{\partial x_j} \ln p \right) \Big|_{\underset{\sim}{x}=\underset{\sim}{\theta}} . \tag{5}$$

Then applying the Schwartz inequality[6] to the right-hand side of Eq. (3) yields

$$(\underset{\sim}{a}, \underset{\sim}{b}) \leq (\underset{\sim}{a}, \underset{\sim}{X}\underset{\sim}{a}) (\underset{\sim}{b}, \underset{\sim}{B}\underset{\sim}{b}) . \tag{6}$$

The matrix $\underset{\sim}{B}$ is positive semi-definite. We assume that it is positive definite, for, if not, we could reduce the dimensionality of our problem. The matrix $\underset{\sim}{B}$ is thus invertible, and we set (since $\underset{\sim}{a}$ and $\underset{\sim}{b}$ are arbitrary vectors) $\underset{\sim}{b} = \underset{\sim}{B}^{-1} \underset{\sim}{a}$. Eq. (6) then becomes

$$(\underset{\sim}{a}, \underset{\sim}{B}^{-1} \underset{\sim}{a})^2 \leq (\underset{\sim}{a}, \underset{\sim}{X}\underset{\sim}{a}) (\underset{\sim}{B}^{-1} \underset{\sim}{a}, \underset{\sim}{a}) . \tag{7}$$

But $\underset{\sim}{B}$ is symetric or self-adjoint, thus $(\underset{\sim}{B}^{-1} \underset{\sim}{a}, \underset{\sim}{a}) = (\underset{\sim}{a}, \underset{\sim}{B}^{-1} \underset{\sim}{a})$ which is greater than zero for any nonzero $\underset{\sim}{a}$. Dividing both sides of inequality (7) by $(\underset{\sim}{B}^{-1}\underset{\sim}{a}, \underset{\sim}{a})$ thus yields the Cramer-Rao inequality

$$(\underset{\sim}{a}, \underset{\sim}{X}\underset{\sim}{a}) \geq (\underset{\sim}{a}, \underset{\sim}{B}^{-1}\underset{\sim}{a}) \tag{8}$$

which holds for an arbitrary vector $\underset{\sim}{a}$.

Now return to the situation described in Sec. I. If we base an estimate on the observations $\underset{\sim}{z}_n$, $n = 1, 2, \ldots, N$, then under the assumptions made on the distribution of the $\underset{\sim}{z}_n$ we have

$$\ln p = \text{const.} + (1/2) \sum_{n=1}^{N} \left[\underset{\sim}{z}_n - \underset{\sim}{f}_n(\underset{\sim}{x}) \right]' \underset{\sim}{\Sigma}_n^{-1} \left[\underset{\sim}{z}_n - \underset{\sim}{f}_n(\underset{\sim}{x}) \right] \tag{9}$$

and

$$\frac{\partial}{\partial x_i} \ln p \Big|_{\underset{\sim}{x}=\underset{\sim}{\theta}} = \sum_{n=1}^{N} \frac{\partial \underset{\sim}{f}_n}{\partial x_i} \Big|_{\underset{\sim}{x}=\underset{\sim}{\theta}} \underset{\sim}{\Sigma}_n^{-1} \left[\underset{\sim}{z}_n - \underset{\sim}{f}_n(\underset{\sim}{\theta}) \right] \tag{10}$$

in which $(\partial \underset{\sim}{f}_n / \partial x_i)$ denotes the vector whose j-th component is $(\partial f_{nj} / \partial x_i)$.

Thus

$$b_{ij} = \sum_{n=1}^{N} \frac{\partial \underset{\sim}{f}_n}{\partial x_i} \Big|_{\underset{\sim}{x}=\underset{\sim}{\theta}} \underset{\sim}{\Sigma}_n^{-1} \frac{\partial \underset{\sim}{f}_n}{\partial x_j} \Big|_{\underset{\sim}{x}=\underset{\sim}{\theta}} . \tag{11}$$

Or, if we let $\underset{\sim}{F}_n$ denote the matrix whose p-q-th element is given by

$$f_{npq} = \frac{\partial f_{nq}}{\partial x_p} \Big|_{\underset{\sim}{x}=\underset{\sim}{\theta}} \tag{12}$$

then the matrix $\underset{\sim}{B}$ appearing in the Cramèr-Rao inequality is

$$\underset{\sim}{B} = \sum_{n=1}^{N} \underset{\sim}{B}_n = \sum_{n=1}^{N} \underset{\sim}{F}_n \underset{\sim}{\Sigma}_n^{-1} \underset{\sim}{F}_n' . \tag{13}$$

This equation, together with Eqs. (12) and (8), gives a lower bound for the mean square error that can be obtained in making an unbiased

estimate of an arbitrary linear combination of parameters, $\Sigma_{j=1}^{R} a_j x_j$, using the observations $\underset{\sim}{z}_n$, n = 1, 2, . . . , N. We will use this bound as a criterion for judging the performance of the estimation method to be discussed in the next section.

III. A RECURSIVE ESTIMATION METHOD

The recursive estimation method to be considered here can be briefly described as follows. After each K-sample times we have a vector observation of the form

$$\underset{\sim}{z}_n = \underset{\sim}{f}_n(\underset{\sim}{x}) + \underset{\sim}{e}_n \qquad n = 1, 2, 3 \ldots$$

in which $\underset{\sim}{f}_n$ is a vector valued nonlinear function of $\underset{\sim}{x}$ whose value uniquely determines the value of $\underset{\sim}{x}$. Suppose, having observed $\underset{\sim}{z}_1$, we carry out a single iteration in the usual iterative or differential correction procedure used to solve the nonlinear problem of finding the maximum likilihood estimate of $\underset{\sim}{x}$ given $\underset{\sim}{z}_1$. Suppose, having observed $\underset{\sim}{z}_2$, we again carry out a single step in this differential correction procedure, basing our calculation on the result of step one and weighting the usual differential correction by the factor 1/2. We then continue in this fashion, basing the single differential correction using the observation $\underset{\sim}{z}_n$ on the estimate resulting from the previous (n - 1) steps and weighting the differential correction by the factor 1/n. The weighting 1/n is used to "average out" the effects of the noise, $\underset{\sim}{e}_n$, as the number of observations becomes large. The idea behind this type of method can perhaps best be understood by reference to the original papers (Ref. 1-3) describing the basic version of this type of method. The interesting point is that although this method can be carried out in real time and is computationally much simpler than a maximum liklihood estimate based simultaneously on all n observations,

the performance of the two methods will be equal, under certain conditions, for large values of n (that is, they will both approach the bound given by the Cramér-Rao inequality).

We will now describe this estimation method more precisely. We denote by $\underset{\sim}{x}_{n+1}$ the estimate of $\underset{\sim}{\theta}$ formed after observing $\underset{\sim}{z}_n$ and we select our initial estimate either arbitrarily or on the basis of some a priori estimate of $\underset{\sim}{\theta}$. We let $\underset{\sim}{M}_n(\underset{\sim}{u})$ denote the matrix whose elements are given by Eqs. (12) with the partial derivatives evaluated at $\underset{\sim}{x} = \underset{\sim}{u}$ instead of $\underset{\sim}{x} = \underset{\sim}{\theta}$. Let

$$\underset{\sim}{y}_n(\underset{\sim}{x}) = \underset{\sim}{F}_n(\underset{\sim}{x})\, \underset{\sim}{\Sigma}_n^{-1}\left[\underset{\sim}{z}_n - \underset{\sim}{f}_n(\underset{\sim}{x})\right] \tag{14}$$

$$\underset{\sim}{B}_n(\underset{\sim}{x}) = \underset{\sim}{F}_n(\underset{\sim}{x})\, \underset{\sim}{\Sigma}_n^{-1}\, \underset{\sim}{F}_n'(\underset{\sim}{x}) \tag{15}$$

and

$$\underset{\sim}{G}_n(\underset{\sim}{x}) = \underset{\sim}{B}_n^{-1}(\underset{\sim}{x}) \,. \tag{16}$$

We assume that there exists some bounded subset, A, of R-dimensional Euclidean space that contains all values of the parameter $\underset{\sim}{x}$ that are regarded a priori as possible (it will be computationally convenient, but not necessary, for A to be a rectangle). We also assume that $\underset{\sim}{\theta}$, the true value of $\underset{\sim}{x}$, lies in the interior of A. Our sequence of estimates $\underset{\sim}{x}_n$, $n = 1, 2, 3, \ldots$, is then determined by the recursive equation

$$\underset{\sim}{x}^*_{n+1} = \underset{\sim}{x}_n + (1/n)\, \underset{\sim}{G}_n(\underset{\sim}{x}_n)\, \underset{\sim}{y}_n(\underset{\sim}{x}_n) \,. \tag{17}$$

If $\underset{\sim}{x}^*_{n+1}$ lies within A, we set $\underset{\sim}{x}_{n+1} = \underset{\sim}{x}^*_{n+1}$, if not we take $\underset{\sim}{x}_{n+1}$ to be the point within A lying closest to $\underset{\sim}{x}^*_{n+1}$.

Note that if the number of measurements or observations made in a group can be picked such that the dimensionality of $\underset{\sim}{x}$ and $\underset{\sim}{z}$ are equal (the dimensionality of $\underset{\sim}{z}$ will always be larger than or equal to that of $\underset{\sim}{x}$) then we have the following simplification

$$\underset{\sim}{G}_n(\underset{\sim}{x})\ \underset{\sim}{y}_n(\underset{\sim}{x}) = \left[\underset{\sim}{F}'_n\ (\underset{\sim}{x})\right]^{-1}\left[\underset{\sim}{z}_n - \underset{\sim}{f}_n(\underset{\sim}{x})\right] .$$

With a mind to making the analysis of Ref. 4 applicable here, we now point out two properties possessed by the quantities $\underset{\sim}{y}_n(\underset{\sim}{x})$ and $\underset{\sim}{G}_n(\underset{\sim}{x})$ by virtue of our definitions and the assumptions made previously in Sec. I.

Property 1. The vector random variables $\underset{\sim}{y}_n(\underset{\sim}{x})$, $n = 1, 2, 3, \ldots$ are statistically independent gaussian random variables (all of whose components have a finite bounded variance).

Now for convenience, let us define

$$\underset{\sim}{m}_n(\underset{\sim}{x}) = E\left\{\underset{\sim}{y}_n(\underset{\sim}{x})\right\} = \underset{\sim}{F}_n(\underset{\sim}{x})\ \underset{\sim}{\Sigma}_n^{-1}\left[\underset{\sim}{f}_n\ (\underset{\sim}{\theta}) - \underset{\sim}{f}_n(\underset{\sim}{x})\right] \tag{18}$$

and note that

$$\underset{\sim}{m}_n(\underset{\sim}{\theta}) = \underset{\sim}{0} . \tag{19}$$

Let $m_{nj}(\underset{\sim}{x})$ denote the j-th component of $\underset{\sim}{m}_n(\underset{\sim}{x})$; then, using Eq. (18) direct evaluation yields

$$\left.\frac{\partial m_{ni}(\underset{\sim}{x})}{\partial x_j}\right|_{\underset{\sim}{x}=\underset{\sim}{\theta}} = -\left[\underset{\sim}{F}_n(\underset{\sim}{x})\ \underset{\sim}{\Sigma}_n^{-1}\ \underset{\sim}{F}'_n\ (\underset{\sim}{x})\right]_{ij}$$

and

$$E\left\{\underset{\sim}{y}'_n\ (\underset{\sim}{\theta})\underset{\sim}{y}_n(\underset{\sim}{\theta})\right\} = \underset{\sim}{F}_n(\underset{\sim}{\theta})\underset{\sim}{\Sigma}_n^{-1}\ \underset{\sim}{F}'_n\ (\underset{\sim}{\theta}) = \underset{\sim}{B}_n .$$

Summarizing these remarks, we have

Property 2. $\underset{\sim}{m}_n(\underset{\sim}{\theta}) = \underset{\sim}{0}$, $\underset{\sim}{G}_n^{-1}\ (\underset{\sim}{\theta}) = \underset{\sim}{B}_n$, and

$$E\left\{y_{ni}(\underset{\sim}{\theta})y_{nj}(\underset{\sim}{\theta})\right\} = b_{nij} = -\left.\frac{\partial m_{ni}(\underset{\sim}{x})}{\partial x_j}\right|_{\underset{\sim}{x}=\underset{\sim}{\theta}} .$$

We now need to make two further assumptions. The first of these does restrict the class of problems for which the recursive estimation method is applicable, but is the type of restriction that must be satisfied by any estimation problem for which an iterative method can be successfully employed to find the maximum value of a nonlinear liklihood function. The second assumption is simply a regularity condition that can always be assumed to hold in practice whenever the first assumption is satisfied. Note that

$$E\left\{\underset{\sim}{z}_n - \underset{\sim}{f}_n(\underset{\sim}{x})\right\} = \underset{\sim}{f}_n(\underset{\sim}{\theta}) - \underset{\sim}{f}_n(\underset{\sim}{x})$$

is zero at $\underset{\sim}{x} = \underset{\sim}{\theta}$, and by our earlier assumption on the uniqueness of the inverse of $\underset{\sim}{f}_n(\underset{\sim}{x})$ is zero only for this value of $\underset{\sim}{x}$. We now further assume that

$$\underset{\sim}{G}_n(\underset{\sim}{x})\underset{\sim}{m}_n(\underset{\sim}{x}) = \underset{\sim}{G}_n(\underset{\sim}{x})\underset{\sim}{F}_n(\underset{\sim}{x})\underset{\sim}{\Sigma}_n^{-1}\left[\underset{\sim}{f}_n(\underset{\sim}{\theta}) - \underset{\sim}{f}_n(\underset{\sim}{x})\right]$$

is also zero only at $\underset{\sim}{x} = \underset{\sim}{\theta}$. In particular, we require

Assumption 1. There exist constants K_o and K_o', $0 \leq K_o \leq K_o' < \infty$ such that

$$K_o \left|\left|\underset{\sim}{x} - \underset{\sim}{\theta}\right|\right|^2 \leq -(\underset{\sim}{x} - \underset{\sim}{\theta})' \underset{\sim}{G}_n(\underset{\sim}{x})\underset{\sim}{m}_n(\underset{\sim}{x}) \leq K_o' \left|\left|\underset{\sim}{x} - \underset{\sim}{\theta}\right|\right|^2$$

for all n and all $\underset{\sim}{x}$ in A, $||\ ||$ denoting the Euclidean norm of a vector.

Our second assumption is

Assumption 2.

$$\underset{\sim}{G}_n(\underset{\sim}{x})\underset{\sim}{m}_n(\underset{\sim}{x}) = -(\underset{\sim}{x} - \underset{\sim}{\theta}) + \underset{\sim}{\zeta}, \quad ||\underset{\sim}{\zeta}|| \leq (1/2)K_1||\underset{\sim}{x} - \underset{\sim}{\theta}||, \quad K_1 < \infty$$

and

$$E\left\{\underset{\sim}{y}_n'(\underset{\sim}{x})\underset{\sim}{G}_n'(\underset{\sim}{x})\underset{\sim}{G}_n(\underset{\sim}{x})\underset{\sim}{y}_n(\underset{\sim}{x})\right\} = \sum_{k=1}^{n} g_{nkk}(\theta) + \tau \quad |\tau| \leq K_2 \quad K_2 < \infty .$$

Note that the essential properties of this assumption have already been implied by Property 2 and Assumption 1; Assumption 2 only bounds the behavior of the remainder terms ζ and τ.

Now let us examine the behavior of the estimates $\underset{\sim}{x}_n$, $n = 1, 2, 3....$ This basic estimation method is considered in Sec. II of Ref. 4; there the character of the observations was not assumed to change with n so that the random variables $\underset{\sim}{y}_n(\underset{\sim}{x})$ were identically distributed for all n and $\underset{\sim}{G}_n(\underset{\sim}{x})$ was independent of n. Nevertheless, that analysis can be easily modified to accomodate the case at hand and show that Properties 1 and 2 and Assumptions 1 and 2 imply the existence of a number N_o such that, for $n \geq N_o$,

$$E\left\{(x_{nj} - \theta_j)^2\right\} \leq (1/n^2) \sum_{k=N_o}^{n} g_{njj}\left[(1 + 2\,k + 1\,k^2)\exp(C/k)\right] + 0(n^{-1+\gamma}) \quad (20)$$

in which C is a bounded constant, γ is some number greater than 0, and $0(n^{-1+\gamma})$ is a term which goes to zero at least as fast as some constant times $n^{-1+\gamma}$. The quantity g_{njj} denotes the j-th diagonal entry of the matrix $\underset{\sim}{G}_n(\underset{\sim}{\theta}) = \underset{\sim}{B}_n^{-1}$.

The quantity inside the brackets in ineq. (20) approaches one for large values of k; thus, for large values of n, we have as an approximation (which can be shown to become exact as $n \longrightarrow \infty$)

$$E\left\{(x_{nj} - \theta_j)^2\right\} \doteq (1/n)\,\overline{g_{jj}(\theta)} = (1/n)\left[(1/n)\sum_{k=N_o}^{n} g_{njj}(\theta)\right] \quad (21)$$

in which we have used the bar to denote a time average. This gives an asymptotic estimate for the mean square error in the j-th component of our parameter vector $\underset{\sim}{x}$. For the purposes of Ref. 4, this was sufficient; here we also wish to find the asymptotic behavior of the covariance matrix of the errors, or, equivalently, to be able to find the asymptotic behavior of the estimation error for any variable which is a linear combination of the original parameters.

To achieve this end, let us consider the change of coordinates

$$\underset{\sim}{v} = \underset{\sim}{C}\underset{\sim}{x} \qquad \underset{\sim}{x} = \underset{\sim}{C}^{-1}\underset{\sim}{v} \ . \tag{22}$$

If we note how the matrices $\underset{\sim}{M}_n(\underset{\sim}{x})$ and $\underset{\sim}{G}_n(\underset{\sim}{x})$ and the vector $\underset{\sim}{y}_n(\underset{\sim}{x})$ were defined, we see that the corresponding quantities in terms of these new v variables will be

$$f_{npq}(\underset{\sim}{v}) = \frac{\partial f_{nq}}{\partial v_p} = \sum_{s=1}^{R} \frac{\partial f_{nq}}{\partial x_s} \frac{\partial x_s}{\partial v_p} = \sum_{s=1}^{R} f_{nsq}(\underset{\sim}{x}) \, (\underset{\sim}{C}^{-1})_{sp} \tag{23}$$

thus

$$\underset{\sim}{F}_n(\underset{\sim}{v}) = (\underset{\sim}{C}^{-1})' \underset{\sim}{M}_n(\underset{\sim}{x}) \tag{24}$$

and

$$\underset{\sim}{y}_n(\underset{\sim}{v}) = \underset{\sim}{F}_n(\underset{\sim}{v}) \underset{\sim}{\Sigma}_n^{-1} \left[\underset{\sim}{z}_n - \underset{\sim}{f}_n(\underset{\sim}{v}) \right] = (\underset{\sim}{C}^{-1})' \underset{\sim}{y}_n(\underset{\sim}{x}) \tag{25}$$

$$\underset{\sim}{B}_n(\underset{\sim}{v}) = \underset{\sim}{F}_n(\underset{\sim}{v}) \underset{\sim}{\Sigma}_n^{-1} \underset{\sim}{F}_n'(\underset{\sim}{v}) = (\underset{\sim}{C}^{-1})' \underset{\sim}{F}_n(\underset{\sim}{x}) \underset{\sim}{\Sigma}_n^{-1} \underset{\sim}{F}_n'(\underset{\sim}{x}) \underset{\sim}{C}^{-1} = (\underset{\sim}{C}^{-1})' \underset{\sim}{B}_n(\underset{\sim}{x}) \underset{\sim}{C}^{-1} \tag{26}$$

$$\underset{\sim}{G}_n(\underset{\sim}{v}) = \underset{\sim}{C} \underset{\sim}{G}_n(\underset{\sim}{x}) \underset{\sim}{C}' \ . \tag{27}$$

Using Eqs. (24-27), one can verify directly that Properties 1 and 2 remain true in the new $\underset{\sim}{v}$ coordinates. Next, consider Assumption 1, letting $\underset{\sim}{\theta}_v = C\underset{\sim}{\theta}$. The quantity in question is

$$\begin{aligned} (\underset{\sim}{v} - \underset{\sim}{\theta}_v)' \underset{\sim}{G}_n(\underset{\sim}{v}) \underset{\sim}{m}_n(\underset{\sim}{v}) &= (\underset{\sim}{x} - \underset{\sim}{\theta})' \underset{\sim}{C}' \underset{\sim}{C} \, \underset{\sim}{G}_n(\underset{\sim}{x}) \, \underset{\sim}{C}' (\underset{\sim}{C}')^{-1} \underset{\sim}{m}_n(\underset{\sim}{x}) \\ &= (\underset{\sim}{x} - \underset{\sim}{\theta}) \, \underset{\sim}{C}' \, \underset{\sim}{C} \underset{\sim}{G}_n(\underset{\sim}{x}) \underset{\sim}{m}_n(\underset{\sim}{x}) \\ &= (\underset{\sim}{C}(\underset{\sim}{x} - \underset{\sim}{\theta}) \ , \ \underset{\sim}{C} \underset{\sim}{G}_n(\underset{\sim}{x}) \underset{\sim}{m}_n(\underset{\sim}{x}) \ ) \ . \end{aligned} \tag{28}$$

Now if $\underset{\sim}{C}$ is a Unitary (or orthogonal) matrix (i.e., represents a rotational transformation), then for any two vectors $\underset{\sim}{a}$ and $\underset{\sim}{b}$ [7]

$$(\underset{\sim}{C}\underset{\sim}{a},\ \underset{\sim}{C}\underset{\sim}{b}) = (\underset{\sim}{a}, \underset{\sim}{b}) \ .$$

We will assume that $\underset{\sim}{C}$ is such a Unitary matrix. Then

$$(\underset{\sim}{v} - \underset{\sim}{\theta}_v)'\underset{\sim}{G}_n(\underset{\sim}{v})\underset{\sim}{m}_n(\underset{\sim}{v}) = (\underset{\sim}{x} - \underset{\sim}{\theta},\ \underset{\sim}{G}_n(\underset{\sim}{x})\underset{\sim}{m}_n(\underset{\sim}{x}))$$

$$= (x - \underset{\sim}{\theta})'\ \underset{\sim}{G}_n(\underset{\sim}{x})\underset{\sim}{m}_n(\underset{\sim}{x}) \qquad (29)$$

so that if Assumption 1 holds in the x set of coordinates it will also hold in the v set of coordinates. Assumption 2 will then also hold in the new set of coordinates (with perhaps different values for the two constants). Thus, any results concerning the behavior of the estimation method as described in the x set of coordinates also pertain when the method is described in the v coordinates. The equivalent of Eq. (22) is

$$E\left\{(v_j - \theta_{vj})^2\right\} \doteq (1/n)\ \overline{g_{jj}(\theta_v)} = (1/n)\left[(1/n)\sum_{k=n_0}^{n-1} g_{kjj}(\theta_v)\right] . \qquad (30)$$

Expressing this equation in the original coordinates, we have

$$\sum_{k,i} c_{jk}\left[\underset{\sim}{X}\right]_{ki} c_{ji} \doteq (1/n)\sum_{k,i} c_{jk}\left[\underset{\sim}{G}_n(\underset{\sim}{\theta})\right]_{ki} c_{ji} \qquad (31)$$

in which $\underset{\sim}{C}$ may be an arbitrary Unitary matrix.

We would like to find the mean square error in the estimate of an arbitrary linear combination of the x variables, $\Sigma_k a_k x_k$, using a corresponding linear combination of our estimates x_{nk}. The error in such an estimate is

$$E\left\{\left[\sum_k a_k(x_{nk} - \theta_k)\right]^2\right\} = \sum_{k,i} a_k\left[X\right]_{ki} a_i \ . \qquad (32)$$

Now the only requirement on a single row of a Unitary matrix is that the sum of the squares of the elements in that row be equal to one. Thus if we scale the a_k's to meet this requirement, the linear combination $\Sigma_k a_k x_k$ can be regarded as some constant times a coordinate of a v vector, $\underset{\sim}{v} = \underset{\sim\sim}{C}\ldots$, in which $\underset{\sim}{C}$ is a Unitary matrix which has one row proportional to the a_k's. Thus Eq. (31) applies and we finally have

$$E\left\{\left(\sum_k a_k x_{nk} - \sum_k a_k \theta_k\right)^2\right\} = (\underset{\sim}{a}, \underset{\sim\sim}{X}\underset{\sim}{a}) \doteq (1/n)\,(\underset{\sim}{a}, \overline{\underset{\sim\sim}{G}_n}\underset{\sim}{a}) \,. \tag{33}$$

We can now compare the performance of the recursive estimation method with the bound given by applying Eq. (26) to the right-hand side of the Cramèr-Rao inequality

$$(\underset{\sim}{a}, \underset{\sim\sim}{X}\underset{\sim}{a}) \geq (\underset{\sim}{a}, \underset{\sim\sim}{B}^{-1}\underset{\sim}{a}) = (\underset{\sim}{a}, (\sum_{k=1}^{n} \underset{\sim\sim}{B}_n)^{-1}\underset{\sim}{a})$$

$$= (1/n)\,(\underset{\sim}{a}, \overline{(\underset{\sim\sim}{G}_n^{-1})}^{-1}\underset{\sim}{a}) \,. \tag{34}$$

Note first that if our measurements are independent of n (as they might be approximately for a high circular orbit) then

$$\overline{(\underset{\sim\sim}{G}_n^{-1})}^{-1} = (\overline{\underset{\sim\sim}{G}_n}^{-1})^{-1} = \overline{\underset{\sim\sim}{G}_n}$$

so that $\overline{\underset{\sim\sim}{G}_n}$, the covariance matrix of the errors in the recursive method, is equal to the covariance matrix of the Cramèr-Rao bound; thus, in this case, the recursive method is asymptotically efficient.

If each of the matrices $\underset{\sim\sim}{G}_n$ does not deviate too much from $\overline{\underset{\sim\sim}{G}_n}$, then $\overline{\underset{\sim\sim}{G}_n}$ is still "close" to the covariance matrix of the Cramèr-Rao bound. To make this precise, let us define the norm of a matrix as

$$||\underset{\sim}{A}|| = \sup_{\underset{\sim}{x}:\|\underset{\sim}{x}\|=1} ||\underset{\sim}{A}\underset{\sim}{x}|| \quad . \tag{35}$$

Note that for a symmetric positive definite matrix $\underset{\sim}{G}$ [8]

$$||\underset{\sim}{G}|| = \max \lambda_G \qquad ||G^{-1}|| = (1/\min \lambda_{\underset{\sim}{G}})$$

in which $\lambda_{\underset{\sim}{G}}$ denotes the eigenvalues of $\underset{\sim}{G}$.

Now the covariance matrix appearing in the Cramer-Rao bound is

$$\underset{\sim}{B}^{-1} = \left[\sum_{k=1}^{n} \underset{\sim}{B}_k\right]^{-1} = \left[\sum_{k=1}^{n} \underset{\sim}{G}_k^{-1}\right]^{-1} \quad . \tag{36}$$

Setting $\underset{\sim}{G}_n = \overline{\underset{\sim}{G}_n} + \underset{\sim}{\Delta G}_n$

$$\underset{\sim}{B}^{-1} = \left[\sum_{k=1}^{n} \overline{\underset{\sim}{G}_k}^{-1} (\underset{\sim}{I} + \overline{\underset{\sim}{G}_k}^{-1} \underset{\sim}{\Delta G}_k)^{-1}\right]^{-1} \quad . \tag{37}$$

We now assume that $||\overline{\underset{\sim}{G}_n}^{-1} \underset{\sim}{\Delta G}_n||$ is small with respect to one (this would be true for example if $(1/\min\lambda_{G_n})(\max\lambda_{\Delta G_n})$ is small with respect to one) so that the inverses inside the inner brackets may be expressed to a good approximation by the first three terms in a Von Neumann series

$$\underset{\sim}{B}^{-1} \doteq \left[\sum_{k=1}^{n} \overline{\underset{\sim}{G}_k}^{-1}\left[\underset{\sim}{I} - \overline{\underset{\sim}{G}_k}^{-1} \underset{\sim}{\Delta G}_k + (\overline{\underset{\sim}{G}_k}^{-1} \underset{\sim}{\Delta G}_k)^2\right]\right]^{-1} \tag{38}$$

But $\overline{\underset{\sim}{G}_n} = (1/n) \sum_{k=1}^{n} \underset{\sim}{G}_k = (1/n) \sum_{k=1}^{n} (\overline{\underset{\sim}{G}_k} + \underset{\sim}{\Delta G}_k)$, thus $\sum_{k=1}^{n} \underset{\sim}{\Delta G}_k = 0$; su stituting this in Eq. (38) and using the first two terms in a Von Neumann expansion of the inverse yields

$$B^{-1} \doteq \overline{\underset{\sim}{G}_n}\left[\sum_{k=1}^{n}\left[\underset{\sim}{I} + \overline{\underset{\sim}{G}_k}^{-1} \underset{\sim}{\Delta G}_k)^2\right]\right]^{-1}$$

$$\doteq (1/n)\, \overline{\underset{\sim}{G}_n}\left[\underset{\sim}{I} - (1/n) \sum_{k=1}^{n} (\overline{\underset{\sim}{G}_k}^{-1} \underset{\sim}{\Delta G}_k)^2\right] . \tag{39}$$

Thus, as long as none of the individual covariance matrices G_n deviates too much from the average ($||\overline{\underset{\sim}{G}_n}^{-1} \Delta\underset{\sim}{G}_n||$ small compared to one), the covariance matrix of the Cramèr-Rao bound does not differ appreciably from the covariance matrix of the errors associated with the recursive estimation method.

REFERENCES

[1] H. Robbins and S. Munro, "A stochastic approximation method," Ann. Math. Stat., Vol. 22, pp. 400-407; 1951.

[2] J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function," Ann. Math. Stat., Vol. 23, pp. 462-466; 1952.

[3] J. Blum, "Multidimensional stochastic approximation methods," Ann. Math. Stat., Vol. 25, pp. 737-744; 1954.

[4] D. Sakrison, "Efficient recursive estimation of the parameters of a covariance function," Internal Technical Memorandum M-76, Electronics Research Laboratory, University of California, Berkeley, California; July 1964.

[5] S. Wilks, "Mathematical statistics," John Wiley and Sons, Inc., New York, pp. 351-352; 1962.

[6] H. Cramer, "Mathematical methods of statistics," Princeton University Press, p. 88; 1946.

[7] G. Shilov, "An introduction to the theory of linear spaces," Prentice-Hall, Inc., Englewood Cliffs, New Jersey, pp. 147-148; 1961.

[8] Ibid., pp. 187-191.